

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-232693

(43)Date of publication of application : 02.09.1998

(51)Int.Cl.

G10L 3/00
G10L 3/00

(21)Application number : 09-161243

(71)Applicant : ATR ONSEI HONYAKU TSUSHIN
KENKYUSHO:KK

(22)Date of filing : 18.06.1997

(72)Inventor : KAWAI ATSUSHI
WAKITA YUMI

(30)Priority

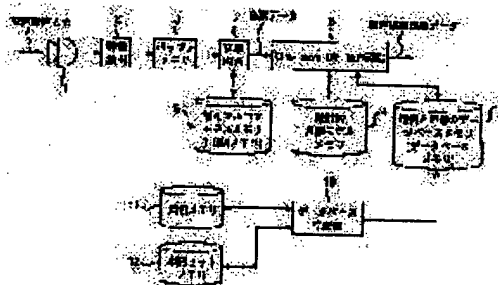
Priority number : 08341084 Priority date : 20.12.1996 Priority country : JP

(54) VOICE RECOGNITION DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a voice recognition device capable of removing an unsuitable erroneous recognition result, outputting a locally and perspective suitable sentence and obtaining a voice recognition rate higher than a usual example.

SOLUTION: This device is provided with a voice recognition part 6 recognizing voice referring to a statistical language model based on a voice signal of a voice of an utterance voice sentence consisting of an inputted word line. The voice recognition part 6 calculates a function value of an unsuitable sentence decisional function for a voice recognition candidate by using a prescribed unsuitable sentence decisional function showing an unsuitable extent for the voice recognition candidate, and when the calculated function value exceeds a threshold value, the voice recognition part 6 removes its voice recognition candidate to voice recognize. The function value is made the value that e.g. the sum of meaning distances answering to examples used in voice recognition processing, and the calculated sum is multiplied by the number of form elements incorporated in the voice recognition candidate becoming the subject of the voice recognition processing, and the value is divided by the number of examples used for the voice recognition processing.



LEGAL STATUS

[Date of request for examination] 18.06.1997

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 2965529

[Date of registration] 13.08.1999

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-232693

(43)公開日 平成10年(1998)9月2日

(51) Int.Cl.⁶

G 1 0 L 3/00

識別記号

5 3 1

5 6 1

FI

G 1 0 L 3/00

5 3 1 P

5 6 1 G

審査請求 有 請求項の数6 OL (全 10 頁)

(21)出願番号 特願平9-161243

(22)出願日 平成9年(1997)6月18日

(31)優先権主張番号 特願平8-341084

(32)優先日 平8(1996)12月20日

(33)優先権主張国 日本 (J P)

特許法第30条第1項適用申請有り 1996年6月24日 社団法人人工知能学会発行の「1996年度人工知能学会全国大会(第10回)論文集」に発表

(71)出願人 593118597

株式会社エイ・ティ・アール音声翻訳通信
研究所

京都府相楽郡精華町大字乾谷小字三平谷 5
番地

(72)発明者 河井 淳

京都府相楽郡精華町大字乾谷小字三平谷5番地 株式会社エイ・ティ・アール音声翻訳通信研究所内

(72)発明者 脇田 由実

京都府相楽郡精華町大字乾谷小字三平谷5番地 株式会社エィ・ティ・アール音声翻訳通信研究所内

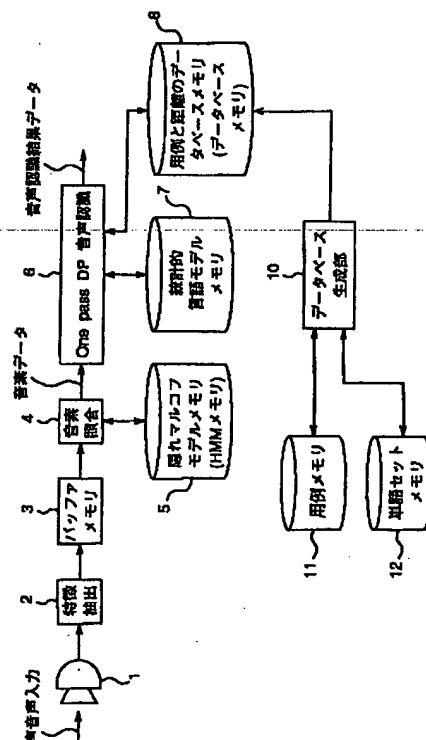
(74)代理人 弁理士 青山 葆 (外2名)

(54) 【発明の名称】 音声認識装置

(57) 【要約】

【課題】 不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供する。

【解決手段】 入力される単語列からなる発声音声文の音声の音声信号に基づいて、統計的言語モデルを参照して音声認識する音声認識部6を備え、音声認識部6は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数を用いて不適格文判定関数の関数値を計算し、計算された関数値がしきい値を超えるとときに、当該音声認識候補を除去して音声認識する。関数値は、例えば、音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値である。



【特許請求の範囲】

【請求項1】 入力される単語列からなる発声音声文の音声の音声信号に基づいて、所定の統計的言語モデルを参照して上記音声に対して音声認識処理を実行する音声認識手段とを備えた音声認識装置において、

上記音声認識手段は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数を用いて不適格文判定関数の関数値を計算し、計算された関数値が所定のしきい値を超えると、当該音声認識候補を除去して音声認識処理を実行することを特徴とする音声認識装置。

【請求項2】 上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値であることを特徴とする請求項1記載の音声認識装置。

【請求項3】 上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値である意味的距離の平均値を計算し、計算された意味的距離の平均値に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値であることを特徴とする請求項1記載の音声認識装置。

【請求項4】 上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値である意味的距離の平均値を計算し、計算された意味的距離の平均値を、所定個の形態素を処理した段階で上記音声認識処理で用いた用例中で所定の複数個以上の形態素を含む用例数で除算した値であることを特徴とする請求項1記載の音声認識装置。

【請求項5】 上記しきい値は、一定値であることを特徴とする請求項1乃至4のうちの1つに記載の音声認識装置。

【請求項6】 上記しきい値は、音声認識処理の対象となる部分文に含まれる形態素の数に依存して変化させることを特徴とする請求項1乃至4のうちの1つに記載の音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、発声音声の音声信号に基づいて、統計的言語モデルを参照して音声認識する音声認識装置に関する。

【0002】

【従来の技術】 連続音声認識装置において、N-gramと呼ばれる統計的手法に基づいた統計的言語モデルが広く使用されている（例えば、従来技術文献1「L.R. Bahl et al. "A Maximum Likelihood Approach to Contin-

uous Speech Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 179-190, 1983年」参照。）。N-gramを用いた連続音声認識装置では、予め、大規模な学習データを用いて、直前のN-1個の単語から次の単語に遷移する遷移確率を学習しておき、音声認識時に、学習した遷移確率を用いて次に接続する単語を予測することにより、音声認識率の向上を計っている。一般に、Nが大きくなるほど次単語の予測精度は向上するが、単語連鎖の種類数が多くなるため、信頼できる遷移確率を得るためには、大量の学習データが必要となる。そこで現状では、Nを2（bigram）又は3（trigram）程度に設定して使用している例が多い。しかしながら、単語のbigramや単語のtrigramを用いた連続音声認識結果を分析してみると、2又は3単語以内の局所的な単語連鎖に自然性はあったとしても、文全体を眺めると、不自然な誤認識文を結果として出力している例が多々見受けられ、より大局的な言語制約が必要であると考える。

【0003】 文脈自由文法などの文法や単語間の依存関係を用いて、より大局的な制約を可能とする言語モデルが提案されている。しかしながら、自然発話文の構造や依存関係の多様性を考えると、規則や依存関係の構築は容易ではないし、処理量も膨大になる。一方、用例主導型のアプローチで文の構文の曖昧性を解消する方法（以下、従来例という。）が従来技術文献2「隅田英一郎ほか、"英語前置語句係り先の用例主導あいまい性解消", 電子情報通信学会論文誌（D-II）, J77-D-II, No 3, pp. 557-565, 1994年3月」において提案されている。この従来例の方法は、コーパスから用例を抽出し、入力文の表現と用例との意味的距離をシソーラスに従って計算し、最終的な意味的距離が最も小さくなる構文を選択する方法であり、対訳決定処理などでもその効果が確認されている（従来技術文献3「古瀬蔵ほか、"経験的知識を活用する変換主導型機械翻訳", 情報処理学会論文誌, Vol. 35, No 3, pp. 414-423, 1994年3月」参照。）。

【0004】

【発明が解決しようとする課題】 しかしながら、従来例の方法を用いる音声認識装置において、例えば、学習した用例に対して不自然な構文を入力すると、どの用例との意味的距離も大きくなってしまい、音声認識率が比較的低いという問題点があった。

【0005】 本発明の目的は以上の問題点を解決し、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することにある。

【0006】

【課題を解決するための手段】 本発明に係る請求項1記

載の音声認識装置は、入力される単語列からなる発声声文の音声の音声信号に基づいて、所定の統計的言語モデルを参照して上記音声に対して音声認識処理を実行する音声認識手段とを備えた音声認識装置において、上記音声認識手段は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数を用いて不適格文判定関数の関数値を計算し、計算された関数値が所定のしきい値を超えるとときに、当該音声認識候補を除去して音声認識処理を実行することを特徴とする。

【0007】また、請求項2記載の音声認識装置は、請求項1記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値であることを特徴とする。さらに、請求項3記載の音声認識装置は、請求項1記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値である意味的距離の平均値を計算し、計算された意味的距離の平均値に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値であることを特徴とする。またさらに、請求項4記載の音声認識装置は、請求項1記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値である意味的距離の平均値を計算し、計算された意味的距離の平均値を、所定個の形態素を処理した段階で上記音声認識処理で用いた用例中で所定の複数個以上の形態素を含む用例数で除算した値であることを特徴とする。

【0008】また、請求項5記載の音声認識装置は、請求項1乃至4のうちの1つに記載の音声認識装置において、上記しきい値は、好ましくは、一定値である。さらに、請求項6記載の音声認識装置は、請求項1乃至4のうちの1つに記載の音声認識装置において、上記しきい値は、好ましくは、音声認識処理の対象となる部分文に含まれる形態素の数の依存して変化させる。

【0009】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。

【0010】図1は、本発明に係る一実施形態の音声認識装置の構成を示すブロック図である。この音声認識装置は、マイクロホン1と、特徴抽出部2と、バッファメモリ3と、入力される発声声文データに基づいて隠れマルコフモデルメモリ（以下、HMMメモリという。）5

Mという。）を参照して音素照合処理を実行して音素データを出力する音素照合部4と、音素照合部4からの音素データに基づいてOne pass DP (Viterbi search) アルゴリズムを用いて統計的言語モデルメモリ7内の統計的言語モデル及び用例と距離のデータベースメモリ（データベースメモリという。）8内の用例と距離のデータベース（以下、データベースという。）を参照して音声認識を実行するOne pass DP 音声認識部（以下、音声認識部という。）6とを備え、上記音声認識部6は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数（詳細後述する数1）を用いて不適格文判定関数の関数値を計算し、計算された関数値が所定のしきい値Fthを超えるとときに、当該音声認識候補を除去して音声認識することを特徴とする。ここで、上記不適格文判定関数の関数値は、好ましくは、上記音声認識候補の構文を決定するために用いた用例に対応する意味的距離の和を計算し、計算された和に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識候補の構文を決定するために用いた用例の数で除算した値である。また、上記しきい値Fthは、好ましくは、一定値、又は、音声認識処理の対象となる部分文に含まれる形態素の数の依存して変化させる。なお、形態素とは、語幹、接頭辞、接尾辞など意味を有する文字系列の最小単位で単語と実質的に同一であるかやや小さい単位である。

【0011】まず、音声認識部6における不適格文検出手法について説明する。N-gramを用いた音声認識処理における誤認識には次の特徴がある。

(a) N個以上の単語の連鎖で判断すると、文法的及び意味的に不適当な単語の組み合わせが存在する。例えば、誤認識例：「電話番号が2107号室ですか」。

(b) 文の構造が大きな単位でまとまらない。すなわち、文法的に規則を適用することができず、局所的にしか判断できない。例えば、誤認識例：「三名様までのえーまでシングルの一泊の……」。

【0012】上記の特徴を持つ誤認識を解決するためには、N-gramよりも、より大局的な立場で、単語間の整合性や構文の適格性を判断する必要がある。一方、用例主導型の音声翻訳手法（従来技術文献2及び従来技術文献4「O.Furuse et al., "Incremental Translation Utilizing Constituent Boundary Patterns", Proceedings of Coling'96, 1996年」参照。）では、用例に基づく翻訳知識を用いて左から右に方向で（left-to-rightに）構文を決定していく方法をとっている。この処理課程で、構文の曖昧性を解消するために、入力文と用例との意味的距離をシソーラス（類語辞書）を用いて計算し、距離の小さい用例に相当する構文を選択する方法をとっている。本発明者は、次の理由により、上記構文決

識を除去するのに整合性が良いと考えられる。

(a) 上記構文決定手法は用例主導型手法であるので、会話文に見られるような従来の文法で処理が困難な構文が容易に処理可能である。

(b) 上記構文決定手法では、構文に基づいて意味的距離を求めているので、隣接しない単語間の整合性を判断できる能力がある。

(c) 音声認識、上記構文決定手法、ともに左から右に方向で (left-to-right) に処理を行なっているので、ある時点までの中間結果を、逐次的に判定できる可能性 10 がある。

【0013】そこで、大局的にみた意味的距離の整合性と解析された構文の適格性で、不適格文を検出する。具体的には次のように判断する。まず、部分文における意味的距離の不整合は、上記の構文決定手法に用いた意味的距離値で判断する。ある部分文の意味的距離の総和が一定値以上になると、その文を誤認識と判断する。次に構文の適格性については次のように考える。一定以上の形態素からなる自然な文であればまとまった構文を持ち、構文の構造はある程度複雑な構造であろうと仮定する。ある部分文に含まれる形態素の数 m の、構文決定のために使用された文脈自由文法の規則又は用例の規則数 (又は用例数) R に対する割合 ($=m/R$) を考える。まとまった構文を持たない部分文は構文構造が階層にならず、よって形態素の数 m に対して、使用された構文規則数 R は少なく、 m/R 値は大きくなる。逆に、構文が複雑になり階層的になるほど、 m/R 値は小さくなる。そこで、次式の不適格文判定関数 $F_{\text{error}}(m)$ を定義する。

【0014】

【数1】

$$F_{\text{error}}(m) = (m/R) \sum_{i=1}^R d(r_i)$$

【0015】ここで、 $d(r_i)$ は複数の用例又は規則 r_i に対応する意味的距離又は類似度距離であり、 m は音声認識処理の対象となる音声認識候補の部分文に含まれる形態素の数であり、 R は音声認識処理を実行するとき音声認識候補の部分文の構文を決定するために用いた用例又は規則の数である。ここで、意味的距離又は類似度距離とは、例えば従来技術文献2のp. 559の

(1) 式で定義され、シソーラスを用いて計算する、入力発声音声文の音声認識候補と用例との間の意味的距離であって、本実施形態においては、音声認識候補の部分文に該当するデータベース内の用例に対する距離を検索

類似度規則

して決定する。ここで、シソーラスとは、概念間の上位下位関係を木構造で表現し、葉に相当する最下位の概念に当該概念をもつ単語を割り当てた辞書を指す。単語間の意味的距離はシソーラス上の概念間の意味的距離によって定義され、概念間の距離はシソーラスにおける最小の共通上位概念の位置に従って0から1までの値に設定される。値0は2つの概念が同じであることを意味し、値1は無関係であることを意味する。また、上記判定関数 $F_{\text{error}}(m)$ は形態素数 m の関数であり、文章の始めから m 番目の形態素までの音声認識候補の部分文を対象に計算される。この判定関数値 $F_{\text{error}}(m)$ が所定のしきい値 F_{th} を越えた場合、音声認識部6は、その音声認識候補の部分文を誤認識結果と判断して、音声認識候補から除去する。なお、上記数1は、好ましくは、 $m \geq 5$ のときに適用することができる。なお、上記数1における規則数 R が0であるときは、当該関数値を1とし、誤認識結果と判断して、音声認識候補から除去する。

【0016】図1の好ましい実施形態においては、データベース生成部10は、用例メモリ11内の用例と、単語セットメモリ12内の単語セットとに基づいて、所定の類似度規則を用いて、データベースを生成して、データベースメモリ8に記憶する。文脈自由文法規則の用例の一例を表1及び表2に示す。また、類似度規則の一例を表3に示す。

【0017】

【表1】用例1

XのY

30

僕の子供
あなたの会社
.....

【0018】

【表2】用例2

XがY

40 僕が先生

.....

【0019】

【表3】

(I) 単語セットの組で生成される文が用例と同じとき、距離=0とする。

(II) 単語セットの組で生成される文が用例と同じ機能単語 (例えば、

を有するとき、距離＝ 10^{-5} とする。

(III) 単語セットの組で生成される文が用例に無い単語同士の時、
距離＝0.5とする。

【0020】日本語処理の音声認識装置における、単語セットS1、S2、S3、S4の一例、並びに、単語セット間の所定の機能単語を用いたときの距離を図2に示す。図2において、例えば、「あなた」(単語セットS1)が「先生」(単語セットS2)のとき、距離が 10^{-5} になり、「あなた」(単語セットS1)の「子供」(単語セットS3)のとき、距離が 10^{-5} になり、「あなた」(単語セットS1)の「会社」(単語セットS4)のとき、距離が 10^{-5} になる。また、「先生」(単語セットS2)の「会社」(単語セットS4)のとき、距離が0.5になる。

【0021】データベース生成部10は、表1及び表2の用例と、表3の類似度規則を用いたときのデータベース生成処理を以下のように行う。各単語セットの組で部分文を生成して、部分文が「あなたの会社」であるときは、距離は0となり、部分文が「私の学校」であるときは、距離は 10^{-5} となり、部分文が「子供が先生」であるときは、距離は0.5となる。このように生成した、部分文の用例と距離とのデータベースは、データベースメモリ8に記憶される。

【0022】さらに、統計的言語モデルは、発声音声文のテキストデータに基づいて、公知の方法により、例えば、単語のbigramの統計的言語モデルを生成して統計的言語モデルメモリ7に記憶する。

【0023】次いで、本実施形態の統計的言語モデルを用いた音声認識装置の構成及び動作について説明する。

【0024】図1において、話者の発声音声はマイクロホン1に入力されて音声信号に変換された後、特徴抽出部2に入力される。特徴抽出部2は、入力された音声信号をA/D変換した後、例えばLPC分析を実行し、対数パワー、16次ケプストラム係数、 Δ 対数パワー及び16次 Δ ケプストラム係数を含む34次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ3を介して音素照合部4に入力される。音素照合部4に接続されるHMMメモリ5内のHMMは、複数の状態と、各状態間の遷移を示す弧から構成され、各弧には状態間の遷移確率と入力コードに対する出力確率を有している。音素照合部4は、入力されたデータに基づいて音素照合処理を実行して音素データを、音声認識部6に出力する。

【0025】統計的言語モデルを予め記憶する統計的言語モデルメモリ7は音声認識部6に接続される。音声認識部6は、統計的言語モデルメモリ7内の統計的言語モデル及びデータベースメモリ8内のデータベースを参照して、所定のOne pass DPアルゴリズムを用

後戻りなしに処理してより高い生起確率の単語を音声認識候補として認識し、当該音声認識候補に対して上記数1を用いて判定関数値Error(m)を計算する。ここで、数1におけるd(r1)は音声認識候補に該当する用例をデータベースより検索して、検索された用例に該当する距離を意味的距離とする。そして、計算された判定関数値Error(m)が所定のしきい値Fthを越えた場合、音声認識部6は、その音声認識候補の部分文を誤認識結果と判断して、音声認識候補から除去する。そして、残った音声認識候補を音声認識結果(文字列データ)と決定して出力する。

【0026】図3は、以上のように構成された日本語処理の音声認識装置の動作を示す動作図であって、入力文と、認識結果文とその構文木とスコアと、構文解析結果文とその構文木とスコアとを示す動作図である。図3

(a)に示すように、「私のエットー学校がね」という入力文の音声が入力されたとき、認識結果文として、図3(b)に示すように、「私の江藤学校がね」が得られたとき、すなわち、「エットー」という間投詞が「江藤」という名詞に誤って認識された場合である。認識結果文における単語間のスコアを図3(b)に示している。さらに、認識結果文に基づいて構文解析したときに、図3(c)に示すように、より小さいスコアに基づいて構文解析結果の構文木が得られ、このときのスコアが得られている。図3(c)における場合を、上記数1に当てはめると、不適格文判定関数の関数値Error(m)は次式のようにになる。

【0027】

【数2】

Error(m)

$= (6/3) (0.5 + 0.5 + 10^{-5})$

$= 2 \times (1.00001)$

$= 2.00002$

【0028】当該例において、不適格文を判定するときのしきい値Fthは、好ましくは、0.6乃至0.7であり、上記数2で計算された関数値 $= 2.00002$ はしきい値Fthを超えているので、それに対応する音声認識候補は音声認識候補から除去される。上記しきい値Fthは、一定値であってもよいし、音声認識処理の対象となる部分文に含まれる形態素数mに依存して変化してもよい。

【0029】以上のように構成された音声認識装置において、特徴抽出部2と、音素照合部4と、音声認識部6と、データベース生成部10とは、例えば、ディジタル計算機などのコンピュータで構成され、バッファメモリ

と、データベースメモリ8とは、例えば、ハードディスクメモリなどの記憶装置で構成される。

【0030】次いで、英語処理の音声認識装置の一例について説明する。英語処理のときの文脈自由文法規則の用例の一例を表4及び表5に示す。また、類似度規則は例えば、表3のものをそのまま使用する。

【0031】

【表4】用例11

X at Y

start at 7:30
leave at 6 p.m.
.....

【0032】

【表5】用例12

Z・X

the train starts
.....

【0033】英語処理の音声認識装置における、単語セットS11(X), S12, S13(Z), S14

(Y)の一例、並びに、単語セット間の所定の機能単語を用いたときの距離を図5に示す。図5において、例えば、「train leaves」のとき距離が 10^{-5} になり、「leave train」のとき距離が0.5になる。また、「leave Kyoto」のとき距離が 10^{-5} になり、「leave at 6 p.m.」のとき距離が 10^{-5} になる。データベース生成部10は、表4及び表5の用例と、表3の類似度規則を用いたときのデータベース生成処理を以下のように行う。各単語セットの組で部分文を生成して、部分文が「the train starts」であるときは、距離は0となり、部分文が「the bus leaves」であるときは、距離は 10^{-5} となり、部分文が「leave yacht」であるときは、距離は0.5となる。このように生成した、部分文の用例と距離とのデータベースは、データベースメモリ8に記憶される。

【0034】図6は、以上のように構成された英語処理の音声認識装置の動作を示す動作図であって、入力文と、認識結果文とその構文木とスコアと、構文解析結果文とその構文木とスコアとを示す動作図である。図6

(a)に示すように、「The bus leaves Kyoto at 11 a.m.」という入力文の音

音声認識及びデータ条件

声が入力されたとき、認識結果文として、図6(b)に示すように、「The bus leaves yacht at 11 a.m.」が得られたとき、すなわち、「Kyoto」という地名の固有名詞が「yacht」という名詞に誤って認識された場合である。認識結果文における単語間のスコアを図6(b)に示している。さらに、認識結果文に基づいて構文解析したときに、図6(c)に示すように、より小さいスコアに基づいて構文解析結果の構文木が得られ、このときのスコアが得られている。図6(c)における場合を、上記数1に当てはめると、不適格文判定関数の関数値F

error(m)は次式のようになる。

【0035】

【数3】

Error(m)

$$= (5/4) (10^{-5} + 0.5 + 0.5 + 10^{-5}) \\ = 1.25 \times (1.00002) \\ = 1.250025$$

【0036】当該例において、上記数3で計算された関数値=1.250025はしきい値Fthを超えているので、それに対応する音声認識候補は音声認識候補から除去される。

【0037】

【実施例】本発明者は、上述の不適格文検出方法を備えた音声認識装置の有効性を評価するために、以下のごとく実験を行った。ここでは、上述の不適格文判定関数Ferrorが、N-gram言語モデルを用いた認識実験における誤認識文と正解文とを区別することが可能かどうかを確認した。具体的には、bi-gramを用いた認識システムによる誤認識結果文と正解文とを対象に不適格文判定関数Errorを算出し、誤認識文と正解文との不適格文判定関数の関数値Errorの違いを考察した。正解文では、形態素の数mが大きい、つまり部分文が長いほど、文構造が複雑になり構造の曖昧性も低くなるので関数値Errorが小さくなり、誤認識文との区別が付きやすくなると想像できる。しかしながら、認識処理の効率化を考えると、なるべく早く、つまり形態素の数mが小さい段階の音声認識候補の部分文に対して不適格判定を行ない、不適格文を誤認識文として結果候補から除去することが好ましい。信頼性の高い関数値Errorを得るための形態素の数mを知るために、誤認識または正解文のm番目の形態素までの音声認識候補の部分文に対して関数値Errorを計算し、形態素の数mを変化させた時の関数値Errorの変化も合わせて調べた。実験における音声認識及びデータ条件を表6に示す。

【0038】

【表6】

音響モデル	不特定話者HM-net, 401状態, 10混合分布
言語モデル	単語のbi-gram
音声認識方式	One-pass DP, N-best探索
bi-gram学習データ	3363文、222954単語
評価データ	学習用データに含まれる44文、4話者

【0039】音声認識処理は、統計的言語モデルに単語のbi-gramを使用し、one-pass DPアルゴリズム、N-best探索型の音声認識システムを用いた。正解文として、表6に示した評価データを用い、誤認識文としては、上記評価データを、表6に示した3種類のN-gramを用いた認識システムで認識し、その結果の誤認識文94文を用いた。図4に、正解文に対する関数値Errorの平均値と最大値、及び誤認識文に対する関数値Errorを、各形態素数m毎に示す。この図4より、次のことがわかる。

(a) 正解文については形態素数mが長くなるほど、関数値Errorの平均値、最大値ともに減少する。

(b) 誤認識文においても同様に、形態素数が長くなるほど関数値Errorは減少する傾向にあるが、その減少の度合いは正解文に比べて少ない。

【0040】このことは、左から右への(left-to-right)の音声認識処理系において、処理した形態素がまだ少ない文の始めの部分では、正解文及び誤認識文の関数値Errorに差がなく、不適格文の検出は困難であるが、処理した形態素数が多くなるほど、正解文と誤認識文との関数値Errorに差が生じるため、上記しきい値Fthを適切に設定することで、不適格文の検出が可能であることを示している。但し、このしきい値Fthは一定値ではなく、形態素数mを変数とする関数値として定義した方がより有効であることがわかる。例えば、図4中の最大値をしきい値Fthとした場合には、このしきい値Fth以上の関数値Errorを示す文章は、各々の形態素数mの処理を行なっている際に、不適格文と判定することができる。このように文の途中結果から誤認識であると判定できた文の割合は、本実験では全誤認識文中47.9% (= 45/94)であった。以上の結果をまとめると、次のようになる。

(a) 本不適格文の検出に用いた(1)入力語句と用例との意味的距離、(2)形態素数に対する規則数で表された文構造の複雑さの2つのパラメータは、不適格文を判定するのに有効なパラメータであり、提案した不適格文判定関数Errorは、不適格文を検出するのに有効であることがわかった。

大きくなるほど、検出性能は上がる。

(c) 不適格文判定関数Errorのしきい値Fthは、形態素数mに依存して変えた方が、より効率良く不適格文を検出できる。

【0041】以上説明したように、本発明によれば、用例との意味的距離を使用することで構文の曖昧性を解消しながら構文を決定していく構文決定手法とを用いて、従来の統計的言語モデルを用いた音声認識の誤認識結果文の不適格性を逐次的に検出する方法を発明した。この方法は、認識結果の部分文に含まれる語句と予め学習された用例との意味的距離と、認識結果の部分文の構文の複雑さを不適格文の判定要因として使用するものである。そして、様々な単語及び品詞のbi-gramを用いた認識システムの結果を対象に、不適格文の検出を行なった結果、誤認識文と正解文との判定のしきい値Fthを適切に設定すれば、誤認識文の約半分を不適格な文として検出可能であることがわかった。

【0042】従って、音声認識部6は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数を用いて不適格文判定関数の関数値を計算し、計算された関数値が所定のしきい値を超えるときに、当該音声認識候補を除去して音声認識するので、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【0043】以上の実施形態においては、不適格文判定関数として数1を用いているが、本発明はこれに限らず、以下に示す数4又は数5の不適格文判定関数を用いてもよい。

【数4】

$$F_{error}'(m) = (m/R) \left\{ \sum_{i=1}^R d(r_i) / R \right\}$$

【数5】

$$F_{error}''(m) = (1/M) \left\{ \sum_{i=1}^R d(r_i) / R \right\}$$

(m) は、数1の不適格文判定関数 $F_{\text{error}}(m)$ に比較して、上記音声認識候補の構文を決定するために用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識候補の構文を決定するために用いた用例の数で除算した値である意味的距離の平均値を計算することを特徴としている。また、数5において、Mは、所定m個の形態素を処理した段階で上記音声認識候補の構文を決定するために用いた用例の規則の中で所定の複数 m_a 個以上の形態素を含む規則数を表し、ここで、mは好ましくは5以上であって、 m_a は好ましくは3である。数5の不適格文判定関数 $F_{\text{error}}'(m)$ は、数3の不適格文判定関数 $F_{\text{error}}(m)$ に比較して、 (m/R) に代えて上記規則数Mの逆数を用いたことを特徴とする。これら数4又は数5の不適格文判定関数を用いて音声認識することにより、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【0045】

【発明の効果】以上詳述したように本発明に係る請求項1記載の音声認識装置によれば、入力される単語列からなる発声音声文の音声の音声信号に基づいて、所定の統計的言語モデルを参照して上記音声に対して音声認識処理を実行する音声認識手段とを備えた音声認識装置において、上記音声認識手段は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数を用いて不適格文判定関数の関数値を計算し、計算された関数値が所定のしきい値を超えるときに、当該音声認識候補を除去して音声認識処理を実行する。従って、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【0046】また、請求項2記載の音声認識装置においては、請求項1記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値である。従って、簡便に上記不適格文判定関数の関数値を計算することができ、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【0047】さらに、請求項3記載の音声認識装置においては、請求項1記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値であ

の平均値に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値である。従って、簡便に上記不適格文判定関数の関数値を計算することができ、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【0048】またさらに、請求項4記載の音声認識装置においては、請求項1記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値である意味的距離の平均値を計算し、計算された意味的距離の平均値を、所定個の形態素を処理した段階で上記音声認識処理で用いた用例中で所定の複数個以上の形態素を含む用例数で除算した値である。従って、簡便に上記不適格文判定関数の関数値を計算することができ、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【0049】さらに、請求項5又は6記載の音声認識装置においては、請求項1乃至4のうちの1つに記載の音声認識装置において、上記しきい値は、好ましくは、一定値、もしくは、音声認識処理の対象となる部分文に含まれる形態素の数に依存して変化させる。従って、より有効的に、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【図面の簡単な説明】

【図1】 本発明に係る一実施形態である音声認識装置のブロック図である。

【図2】 図1の音声認識装置における日本語の単語セットと距離との関係を示す図である。

【図3】 図1の音声認識装置の日本語処理の動作を示す動作図であって、入力文と、認識結果文とその構文木とスコアと、構文解析結果文とその構文木とスコアとを示す動作図である。

【図4】 図1の音声認識装置のシミュレーション結果であって、入力された形態素の数に対する判定関数値 F_{error} を示すグラフである。

【図5】 図1の音声認識装置における英語の単語セットと距離との関係を示す図である。

【図6】 図1の音声認識装置の英語処理の動作を示す動作図であって、入力文と、認識結果文とその構文木とスコアと、構文解析結果文とその構文木とスコアとを示す動作図である。

15

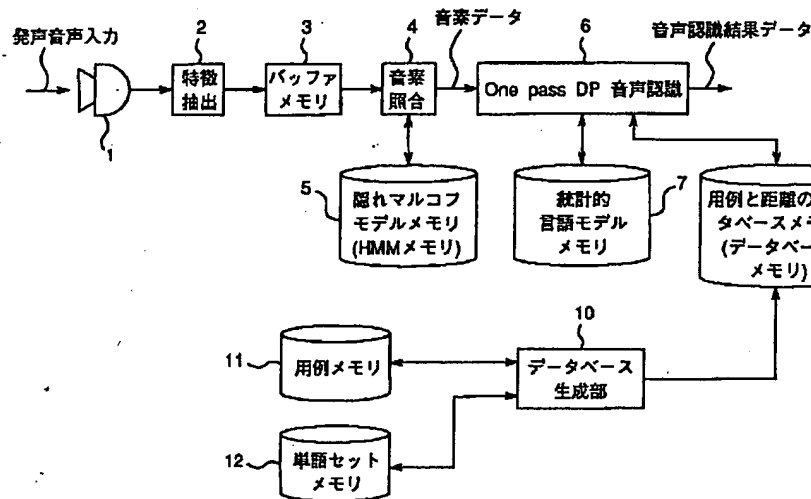
16

- 1…マイクロホン、
 2…特徴抽出部、
 3…バッファメモリ、
 4…音素照合部、
 5…隠れマルコフモデルメモリ (HMMメモリ)、
 6…One pass DP音声認識部、

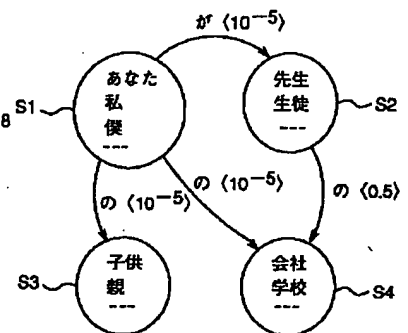
- 7…統計的言語モデルメモリ、
 8…用例と距離のデータベースメモリ (データベースメモリ)、
 10…データベース生成部、
 11…用例メモリ、
 12…単語セットメモリ。

【図1】

【図2】

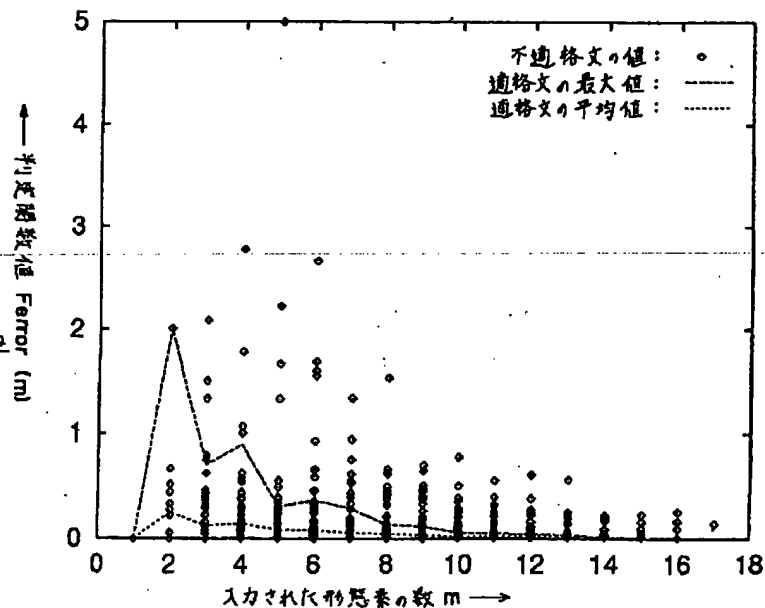
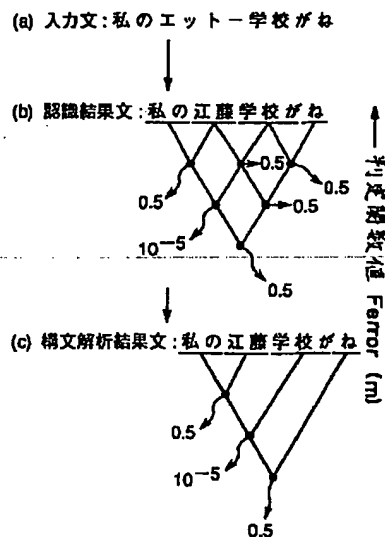


日本語の単語セットと距離

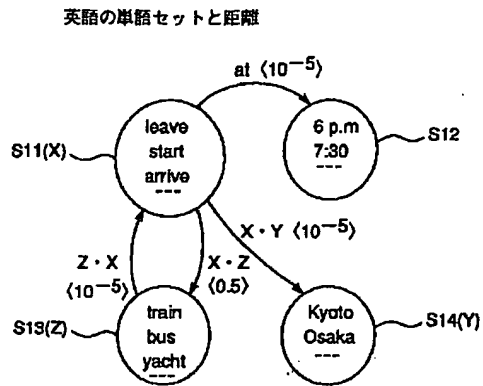


【図3】

【図4】



【図5】



【図6】

